

LIBRARY OF THE
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

510.84

Il6r

no.493-498

cop.2



510.84
IL6 v
no. 496
cop. 2

math

UIUCDCS-R-72-496

COO-2118-0028

Methodological Aspects
of Scene Segmentation

By
Peter Raulefs

August 1972



DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN · URBANA, ILLINOIS

THE LIBRARY OF THE

SEP 5 1972

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN



Digitized by the Internet Archive
in 2013

<http://archive.org/details/methodologicalas496raul>

COO-2118-0028

UIUCDCS-R-72-496

Methodological Aspects
of Scene Segmentation

By
Peter Raulefs

Department of Computer Science
University of Illinois
Urbana, Illinois 61801

This work was supported in part by Contract AT(11-1)-2118
with the U.S. Atomic Energy Commission.

ACKNOWLEDGMENT

The author would like to express his gratitude to Professor Bruce H. McCormick who introduced him into the area of image processing and contributed to this thesis in numerous discussions and suggestions.

The author is also grateful to the members of the Illiac III group for pleasant and helpful cooperation. In particular, he acknowledges the help of Mrs. Judy Arter for typing and improving this thesis through editing, and he appreciates the excellent drawings done by Mr. Stanley Zundo.

TABLE OF CONTENTS

CHAPTER	Page
1 INTRODUCTION	1
2 SCENE SEGMENTATION	3
2.1 Property Space Representation	3
2.1.1 Local Representation of Pictures	3
2.1.2 Definition and Representation of Regions	5
2.1.3 Criterion for Partitioning Pictures into Regions	9
2.1.4 Density and Entropy of Sets in the Property Space	10
2.1.5 Formal Description of Scene Segmentation into Regions	13
2.1.6 Metrics in the Property Space	14
2.2 The Clustering Approach to Scene Segmentation	18
2.2.1 A Formal Concept of Clustering	19
2.2.2 Application of Clustering to Scene Segmentation	23
2.2.3 Review of Some Clustering Techniques	25
2.2.4 A Combined Clustering Technique for Scene Segmentation	28
3 STRUCTURAL ANALYSIS OF SCENES	33
3.1 Data Representation of Composite Scenes	33
3.2 Synthesis of Composite Objects in Scenes	35
3.2.1 Synthesis of Composites Using Graph Transformation Rules	36
3.2.2 Synthesis of Composites Using Cartesian Covers	36
3.2.3 Matching Models to Scenes	38
4 CONCLUSIONS	40
REFERENCES	41

LIST OF FIGURES

Figure	Page
1a (6,3) - Regular Tessellation of the Plane	7
1b (4,4) - Regular Tessellation of the Plane	7
1c (3,6) - Tessellation of the Plane	8
2a (4,4) - Tessellation Which is 4-Imbedded in Another (4,4) - Tessellation	21
2b Cohesions of Simple Regions	21
3 Cohesions of Different Regions in a Scene	22

CHAPTER 1

INTRODUCTION

We consider scene analysis to consist of two major parts: picture preprocessing and structural analysis.

Scenes are viewed as being pictorial representations of an assembly of more or less well known objects. In image processing research, a scene is usually given as a digitized picture constituting the output of a scanning device. The raster points of a digitized picture are referred to as picture points. The preprocessing phase consists of partitioning the picture into sets of picture points. Subsequently, these sets have to be interpreted as representations of certain objects (structural analysis).

The basic idea of this paper is that the representation of objects in a scene is given by sets of picture points such that the local properties of points within such a set are more similar to each other than those belonging to different sets. Such a set of picture points will be called a region. The idea of partitioning scenes into regions has proven to yield a convenient data structure for subsequent processing techniques (see [BF 70], [GUZ 70], [BP 70], PR 70] and Sect. 3 of this paper).

Although this approach to region definition is closely related to the concept of clustering, a precise definition has not yet been given. It is shown in Sect. 2.1 that the information-theoretical quantity entropy is useful in characterizing a region. In Sect. 2.2, entropy is used to define the cohesion of a region, and this term gives a tool to analyze clustering methods and to specify an algorithm giving an optimal scene segmentation.

After having partitioned a picture into regions, each region must be interpreted as a part of a known object. This requires a model of how

objects may be represented and which relations between these representations must hold. Then, recognition can be viewed as finding an optimal match of regions to a collection of model objects. Although we do not pursue a solution to this problem in this paper, regions are used to define a data structure which is applied to approaches to scene analysis briefly reviewed in Sect. 3.

SCENE SEGMENTATION

2.1 Property Space Representation

2.1.1 Local Representation of Pictures

A picture P is considered to be given as a finite set of points in the plane that are associated with local attributes.

Definition 2.1 Let P be a finite set of discrete points in the plane.

A local attribute, A_t , is a function

$$A_t: \begin{cases} P \rightarrow V_t = V'_t \cup \{*\} \\ p_m \rightarrow A_t(p_m) \end{cases}$$

mapping each point $p_m \in P$ into the value set, V_t , where V_t consists of a finite set V'_t of real numbers and the don't care value, $*$.

Remarks:

1. An attribute A_t assumes the don't care value, $A_t(p_m) = *$, at a point $p_m \in P$ whenever no real number is specified as the value of A_t at p_m .
2. We assume that the attributes A_t are always specified so that the values in V'_t are naturally ordered real numbers, but we do not assume any ordering between the elements in V'_t and $*$.
3. We do not assume that a distance function is known in any of the sets, V_t .
4. With reference to some coordinate systems in the plane, we adopt the convention that for any point, $p \in P$, with coordinates x_1 and x_2 , the attributes A_1 and A_2 are defined by $A_i(p) = x_i$ ($i=1,2$).

With these notions we consider a picture, P , to be represented as a finite set of points in the plane such that each picture point, p , is associated with a N -dimensional vector $(A_1(p), \dots, A_N(p))$ of attribute values.

Definition 2.2 Let $A = \{A_1, \dots, A_N\}$ be a set of local attributes. The property space $X(A)$ is defined as the set

$$X(A) = \prod_{t=1}^N V_t .$$

Definition 2.3 Let P be a picture consisting of M points $p_1, \dots, p_M$. The property space representation of P is defined to be the set $P = \{\underline{x}_1, \dots, \underline{x}_M\}$ of M points in $X(A)$:

$$P = \{ \underline{x}_m \mid \text{for } 1 \leq m \leq M: \underline{x}_m = (A_1(p_m), \dots, A_N(p_m)) \} .$$

Remark: To obtain a convenient representation of pictures, the don't care value, $*$, may be associated with a real number, x_* , which is not contained in any of the sets $V_1, \dots, V_N$. We denote by $X'(A) = \prod_{t=1}^N V'_t$ the subset of $X(A)$ that contains the N -tuples of all specified attribute values.

Example: To illustrate the property space representation of a picture, we consider the gray value of picture points as an attribute A_3 . The gray value is defined by quantizing the possible intensities of light associated with picture points and assigning a number to each intensity interval thus obtained. The corresponding property space,

$X(A_1, A_2, A_3)$, is three-dimensional, where two dimensions are used to describe the coordinates and the third to give the gray values of picture points.

As the gray values of single picture points are often influenced by noise effects, attributes obtained, for example, by averaging the gray values over neighborhoods, may be considered.

2.1.2. Definition and Representation of Regions

The property space representation can be considered as a local description of pictures. A step towards a more global description is the formation of regions. This is motivated by the following considerations:

1. From evaluating the importance of various parts of pictures for human perception of form, F. Attneave concluded that the "information contents" of a black-white picture is concentrated at points where the gradient of the density distribution of black points is large ([ATT54], [AA66]). It is suggested in [ATT54] that reducing a picture to a contour-line drawing reveals some redundancy among points circumscribed by closed contour-lines. This interpretation of experimental results in terms of information theory has been seriously questioned in [GC66]. But we can safely infer that the representation of pictures as an arrangement of regions (1) leads to a higher stage of conceptualization, and (2) reduces the number of code words necessary to represent a picture for automated picture processing.

2. Scene segmentation into regions is a convenient step of generalization allowing the application of graph processing techniques. A similar data structure has been used in [BF70], [GUZ70], [BP70], and [PR70].

A. Regions in Discrete Point Sets

Definition 2.4 Let P be a set of M points in the plane such that the coordinates $\underline{x}^m = (x_1^m, x_2^m)$ of each point in P are defined with respect to some coordinate system.

For any $\epsilon > 0$, a subset $U \subseteq P$ is called ϵ -chain-connected if for any pair $\underline{x}, \underline{x}' \in U$, there is a chain $\underline{t}^1, \dots, \underline{t}^k \in U$ such that:

- (i) $\underline{x} = \underline{t}^1$ and $\underline{x}' = \underline{t}^k$;
- (ii) for $j = 2, \dots, k$: $||\underline{t}^{j-1} - \underline{t}^j|| \leq \epsilon$,
 where $||\underline{x}|| = \sqrt{x_1^2 + x_2^2}$ denotes the Euclidean norm of a vector, $\underline{x} = (x_1, x_2)$.

Definition 2.5 An ϵ -chain-connected set of picture points in a picture, P , is called an ϵ -region in P for any $\epsilon > 0$.

B. Regions in Cellular Images

In many applications, the acquisition of pictorial data is done by scanning pictures at all points of a pre-selected regular grid. A systematic method of defining a grid is to tessellate the plane into (q) regular p -gons as discussed in [GR71]. It can be shown (cf. [GR71]) that there are only three regular tessellations of the plane for $(p, q) = (6, 3)$, $(4, 4)$ and $(3, 6)$. Then, a regular grid, G , is obtained by placing grid points at the barycenters of all p -gons. Each p -gon is called a cell.

Definition 2.6 Let P be a set of M points of a regular grid, G , obtained by a regular tessellation of the plane into (q) regular p -gons (cells). A set, $U \subseteq P$, is G -chain-connected if for any two points $p, p' \in U$, there is a chain $q_1, \dots, q_k \in U$ such that:

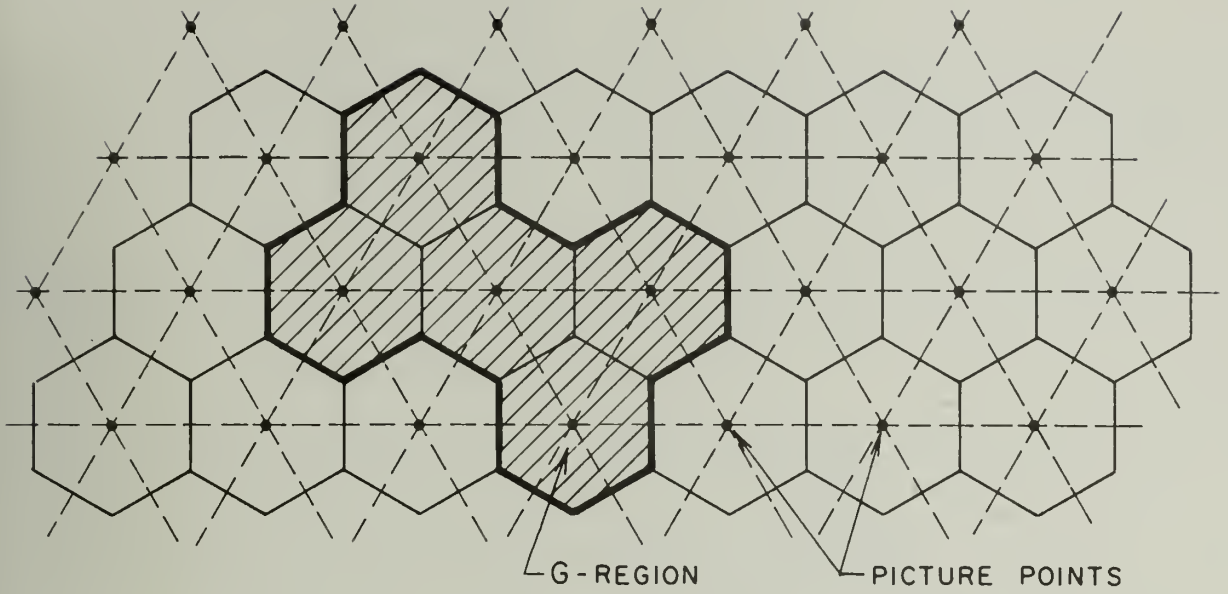


Figure 1a (6,3) - Regular Tessellation of the Plane

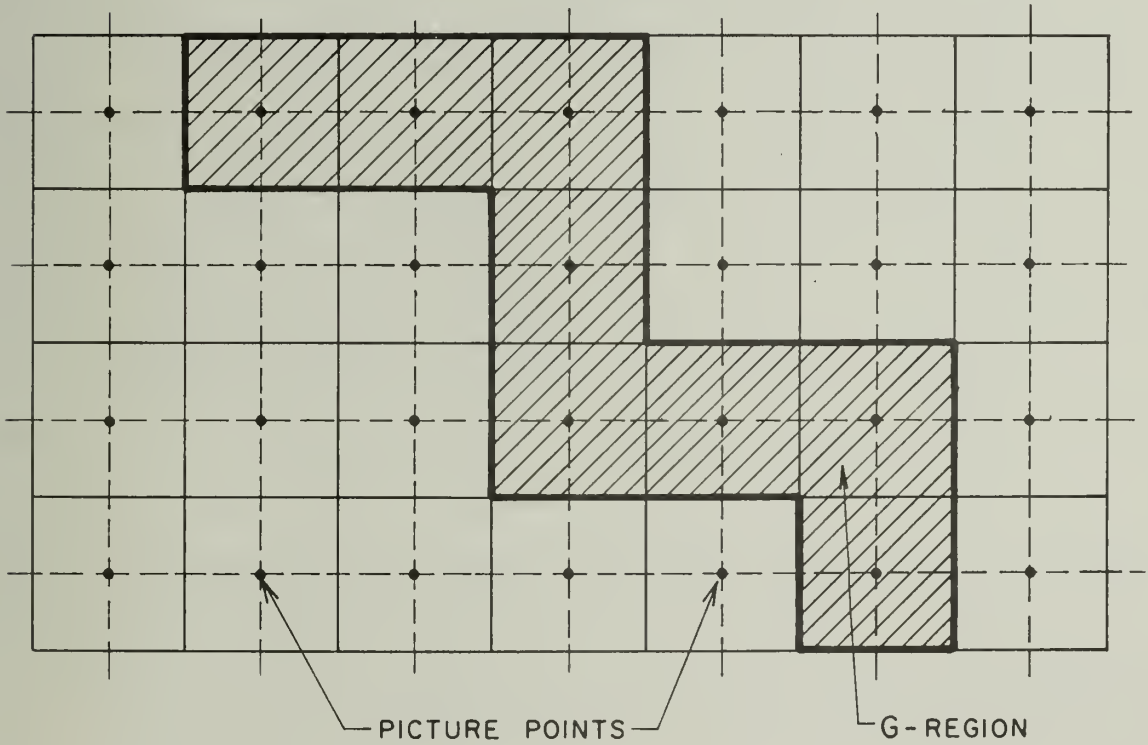


Figure 1b (4,4) - Regular Tessellation of the Plane

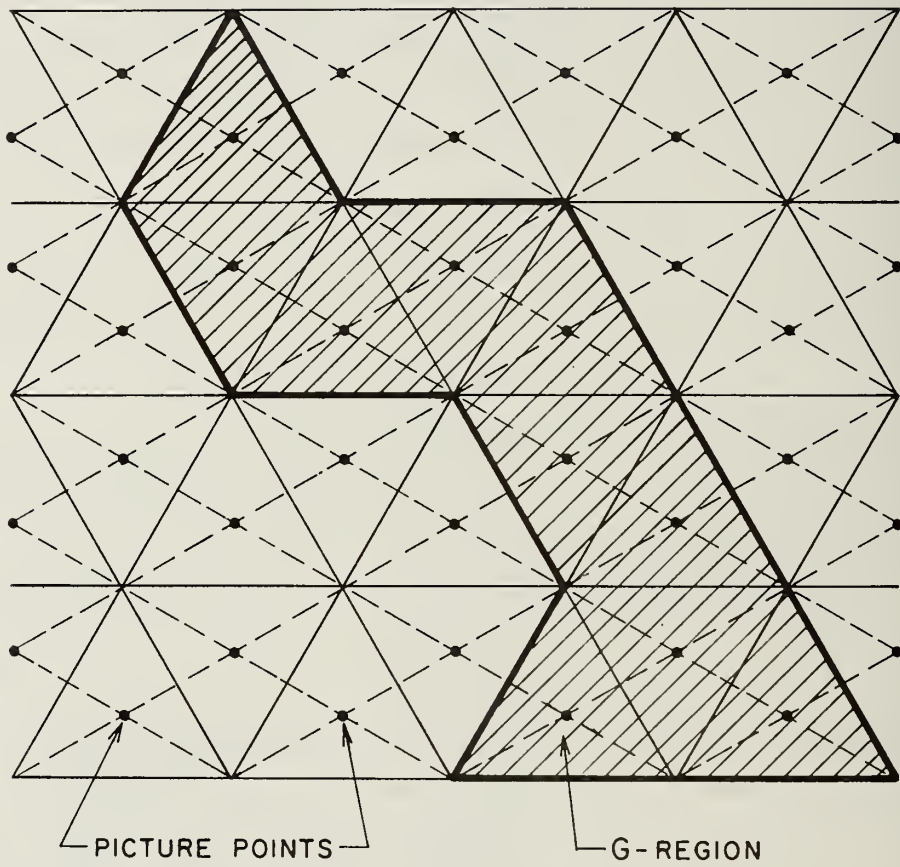


Figure 1c (3,6) - Tessellation of the Plane

- (i) $p = q_1$ and $p' = q_k$;
- (ii) for $j = 2, \dots, k$: $c(q_j)$ and $c(q_{j-1})$ are edge-adjacent or vertex-adjacent, where $c(q)$ is the cell containing the point q .

Definition 2.7 A G-chain-connected set of picture points in a picture, P , is called a G-region in P .

Regular tessellations and G-regions are illustrated in Figures 1a-c.

2.1.3 Criterion for Partitioning Pictures Into Regions

Early approaches to region finding, such as [ROB65], employ two steps: (1) edge finding; and (2) fitting the line drawing obtained after (1) to a model. In this section, we are interested in finding regions without regard to a model, thus assuming an early stage of image preprocessing.

A different method, proposed in [MP67], bypasses the edge finding phase by directly forming regions as the union of squares with approximately the same gray value. In [BF70], regions are constructed by joining "atomic regions" according to two heuristic rules. In our notation, these "atomic regions" are chain-connected sets of picture points whose properties satisfy certain equivalence relations, e.g. equality of gray values. The method of [BF70] is based on the notion of the "strength of boundary", which is defined as the sum of the distances (in the plane) of all adjacent points separated by the boundary. The heuristics are then aimed at merging regions if the boundaries separating them are weak.

The region finding mechanism of these approaches can be generalized to the following scheme:

1. Define a similarity measure for (ϵ -resp.G-) chain-connected

regions represented in the property space.

Remark: A set consisting of one picture point only is always chain-connected.

2. Assuming some threshold value θ , merge any two chain-connected regions to a single region if their similarity is larger than θ .

Remark: Two regions S_1 and S_2 are (ϵ -resp. G-)chain-connected if any pair $(p_1, p_2) \in S_1 \times S_2$ is chain-connected.

The definition of a similarity measure depends on defining a metric in the property space. Then, the threshold value, θ , can be specified as that distance in the property space up to which regions are considered to be similar. By definition of the property space representation of pictures, the planar coordinates are also coordinates in the property space X . Consequently, the above partitioning scheme yields a segmentation into regions such that the density distribution of points in X is more uniform in each region than in the entire picture. This can be further generalized to the following criterion:

[Region Criterion] An optimal partition of a picture into regions is given by:

1. minimizing the average intra-set dispersion within regions, and
2. maximizing the average inter-set dispersion between regions.

To obtain a formal and precise statement of this criterion, the concept of a density distribution of points in the property space will be introduced in the next section. With this notation, we shall see that the [Region Criterion] can very well be interpreted in terms of information theory.

2.1.4 Density and Entropy of Sets in the Property Space

As the property space, $X(A)$, is discrete and finite, the definition

of a density distribution of points in subsets of $X(A)$ depends critically on the neighborhoods in which densities are evaluated. These densities will be applied by interpreting them as probability densities. Therefore, we have to establish non-overlapping neighborhoods to define normalized densities. The lattices considered in crystallography are a useful notion for this purpose, as lattices are generated by applying elements of the translation group to unit cells (see, for example, [KIT63]).

The property space $X(A)$ is imbedded in the N -dimensional vector space, R^N , of N -tuples of real numbers. As points in $X(A)$ may be distributed quite irregularly, we establish neighborhoods by partitioning R^N into N -dimensional cubes with sides of length $\hat{\ell}$.

Definition 2.8 Let $\hat{x}^1, \dots, \hat{x}^N$ be an ordered set of orthogonal vectors in R^N spanning an N -dimensional cube with sides of length $\hat{\ell}$.

The set $C(\underline{x}^0) = \{\underline{x} \in R^N \mid \exists \lambda_1, \dots, \lambda_N \in [0,1[: \underline{x} = \underline{x}^0 + \sum_{n=1}^N \lambda_n \hat{x}^n\}$ is called a unit cell with origin $\underline{x}^0$.

We consider translations of the unit cell described by translation operators T_λ^n :

$$T_\lambda^n[C(\underline{x}^0)] = \{\underline{x}' \in R^N \mid \exists \underline{x} \in C(\underline{x}^0): \underline{x}' = \underline{x}^0 + \lambda \hat{x}^n\}$$

for $1 \leq n \leq N$ and $\lambda = 0, \pm 1, \pm 2, \dots$

Definition 2.9 A standard tessellation, $T(\underline{x}^0, \hat{\ell})$, of R^N into unit cells is defined to be the set of all unit cells such that:

(i) $C(\underline{x}^0)$ is a unit cell with origin $\underline{x}^0$ and sides of length $\hat{\ell}$;

(ii) if $C(\underline{x}') \in T(\underline{x}^0, \hat{\ell})$ then there are translations

$$T_{\lambda_1}^{n_1}, \dots, T_{\lambda_r}^{n_r} \text{ such that}$$

$$T_{\lambda_1}^{n_1}(T_{\lambda_2}^{n_2}(\dots(T_{\lambda_r}^{n_r}(C(\underline{x}^0))\dots)) = C(\underline{x}') .$$

Remarks:

1. By Definition 2.8 all unit cells contained in a standard tessellation of R^N do not overlap.
2. Whenever the length, $\hat{\ell}$, is not relevant, a standard tessellation, $T(\underline{x}^0, \hat{\ell})$ will be referred to as $T(\underline{x}^0)$.

Definition 2.10 Let S be a set of M points in $X'(A)$ and $T(\underline{x}^0)$ a standard tessellation. The normalized density, $\rho_S(\underline{x})$, of S at a point $\underline{x} \in S$ is defined by

$$\rho_S(\underline{x}) = \frac{1}{M} |S \cap C|, \quad \text{where } C \in T(\underline{x}^0) \text{ is a unit cell with } \underline{x} \in C.$$

The normalized density has the properties $\sum_{\underline{x} \in S} \rho_S(\underline{x}) = 1$ and $0 \leq \rho_S(\underline{x}) \leq 1$ for all $\underline{x} \in S$. Therefore, we can utilize ρ_S to define the entropy of the set S .

Definition 2.11 The entropy $H(S)$ of a set S of M points in $X'(A)$ for a given standard tessellation, $T(\underline{x}^0)$ is defined by

$$H(S) = - \sum_{\underline{x} \in S} \rho_S(\underline{x}) \log \rho_S(\underline{x}) .$$

The entropy, $H(S)$, has the usual properties discussed in information theory and thermodynamics: $0 \leq H(S) \leq \log M$, and $H(S) = \log M$ iff $\forall \underline{x} \in S: \rho_S(\underline{x}) = \text{Const.}$

and H is a U -convex function (cf. [GAL68]).

2.1.5 Formal Description of Scene Segmentation into Regions

The ideas developed in the previous section can now be applied to the segmentation of scenes into regions.

By definitions 2.5 and 2.7, a region, S , is an ϵ -resp. G -chain-connected point set in the plane of picture points. We denote the property space representation of a region S by S^X . Clearly, S^X is not necessarily a chain-connected set with respect to a standard tessellation of $X(A)$, i.e. the unit cells containing points of S^X may not be edge- or vertex-adjacent. However, the projection of S^X into $V_1' \times V_2'$ is ϵ -resp. G -chain-connected.

The intra-set dispersion of points in S^X is measured by the negentropy $-H(S^X)$. If the density distribution of points in S^X is uniform, $H(S^X)$ attains its maximum, $H(S^X) = \log |S^X|$. Consequently, the problem of finding regions by minimizing the average intra-set dispersion of points is equivalent to maximizing the average entropy of regions.

To compare the dispersion of points of different regions in $X(A)$, we have to define a measure which is independent of the area occupied by particular regions in the plane.

Definition 2.12 The normalized entropy, $\hat{H}(S^X)$, of a set S^X of $M > 1$ points in $X(A)$ is given as

$$\hat{H}(S^X) = \frac{1}{\log M} H(S^X)$$

For $M=1$: $\hat{H}(S^X)=0$.

Remark: The normalized entropy of a set, S^X , in $X(A)$ satisfies $0 \leq H(S^X) \leq 1$. For a given segmentation, $S = \{S_1, \dots, S_L\}$, of a picture into L regions, the average intra-set dispersion is measured by means of

$$\bar{H}(S) = \frac{1}{L} \sum_{\ell=1}^L H(S_{\ell}^X) .$$

Minimizing the average intra-set dispersion is equivalent to maximizing $\bar{H}(S)$ over all possible partitions of S into regions.

2.1.6 Metrics in the Property Space

A. Objectives

The decomposition of scenes into regions depends decisively on how the distance of any two points in the property space is evaluated. We will restrict ourselves to considering weighted Euclidean metrics, i.e., for any two points $\underline{x}, \underline{y} \in X$, the distance is given as

$$d(\underline{x}, \underline{y}) = \left\{ \sum_{n=1}^N [\omega_n (x_n - y_n)]^2 \right\}^{1/2}$$

where $\omega_n \geq 0$ is the weight for the n -th coordinate. This definition of a distance is equivalent to multiplying each coordinate of vectors in X with a particular weight (see [SEB62]). Hence, $(\omega_i \Delta_i) \gg (\omega_j \Delta_j)$ can be interpreted as regarding the i -th coordinate to be more important than the j -th coordinate, where $\Delta_{\ell} = \max\{x_{\ell}\} - \min\{x_{\ell}\}$.

The necessity of specifying the relative importance of each coordinate with respect to all others can be seen by applying the Theorem of the

Ugly Duckling ([WAT69a], [WAT69b]) to vectors in the property space, X :

Let each coordinate of vectors in X represent a predicate with values 0 or 1, and let all possible predicates be represented as coordinate axes of X . Then, with each predicate, its negation is also a coordinate axis of X . The Theorem of the Ugly Duckling states that any pair of two vectors in X are as similar to each other as any other pair of two vectors, where the similarity is given as the number of corresponding coordinates which are equal. This "similarity" is identical to the Hamming distance of binary sequences. The Theorem of the Ugly Duckling still holds whenever any finite number of values can be attained by each coordinate and the similarity is defined in the same way as the number of equal, corresponding coordinates. As predicates with a finite number of discrete values are the properties defined in Section 2.1.1, we conclude that to introduce dissimilarities between vectors in X , the properties have to be selected and weighted.

The procedure of attaching weights to properties is usually referred to as feature selection. As the relative importance of features is not known a priori, we have to give a training algorithm to determine the weights. Although the problem of feature selection has been extensively investigated in the literature, these studies do not lead to efficient algorithms decomposing scenes into regions. An approach to feature selection using entropy maximization is given in [TOU69]. This method has characteristics typical to many other feature selection techniques: It is assumed that (1) the number of classes to be recognized is known and that (2) the components of feature (i.e. property) vectors are normally distributed random variables. Both assumptions are usually not applicable to picture segmentation.

To obtain an algorithm that computes a metric from a given training set of examples, we consider the following system;

1. A training set, $\{P_1, \dots, P_t, \dots, P_T\}$, of examples is given, consisting of several pictures already decomposed into regions.
2. The distance between two points $\underline{x}, \underline{y} \in X$ is given by the weighted Euclidean distance

$$d(\underline{x}, \underline{y}) = \left\{ \sum_{n=1}^N \omega_n^2 (x_n - y_n)^2 \right\}^{1/2}$$

with the constraint $\prod_{n=1}^N \omega_n = 1$. This constraint guarantees that metrics derived from pictures differing only by a linear transformation (e.g. shrinkage) will be equal (cf. [SEB62]).

Remark: Without the constraint $\prod_{n=1}^N \omega_n = 1$, d is not necessarily a metric, but only a pseudo-metric (see [BIR67]), as $d(\underline{x}, \underline{y}) = 0$ is possible even when $\underline{x} \neq \underline{y}$ for some weight vectors $\underline{\omega} = \{\omega_1, \dots, \omega_N\}$.

3. The weights $\{\omega_1, \dots, \omega_N\}$ are to be adjusted so that $P_1, \dots, P_T$ satisfy the [Region Criterion] of Section 2.1.5.

According to the [Region Criterion], $\bar{H}(S)$ is to be maximized for the training set of sample pictures. To obtain a maximized $\bar{H}(S)$, we have to adjust the weight vector, $\underline{\omega}$, so that the density is distributed as uniformly as possible within each region of $P_1, \dots, P_N$.

B. Algorithm to Determine a Pseudo-Metric in the Property Space

Maximization Problem:

We assume that a standard tessellation of X and a training set, $S = \{S_1, \dots, S_\ell, \dots, S_L\}$, of L regions is given, where the ℓ -th region, S_ℓ , consists out of M_ℓ cell groups. At this point, the coordinate axes

of X have some arbitrary scaling and we assume a weight vector $\underline{\omega}$ to be $\underline{\omega} = (1, \dots, 1)$. Any change of $\underline{\omega} = (\omega_1, \dots, \omega_N)$ changes a coordinate value, x_n , to $\omega_n x_n$. If the unit cell, C , of the standard tessellation was initially given by some origin, $\underline{x}_0 = (x_{01}, \dots, x_{0N})$ and $\Delta \underline{x}' = (\Delta x_1, \dots, \Delta x_N)$, this will change to $\underline{x}' = (\omega_1 x_{01}, \dots, \omega_N x_{0N})$ and $\Delta \underline{x}' = (\omega_1 \Delta x_1, \dots, \omega_N \Delta x_N)$. The number of elements of S_ℓ contained in the unit cell that contains a vector, $\underline{x} \in X$, is denoted by $\Delta n(\underline{x}, \Delta \underline{x})$. Thus, the normalized density, $\rho(S_\ell, \underline{x})$, is given by

$$\rho(S_\ell, \underline{x}) = \frac{1}{M_\ell} \cdot \frac{\Delta n(\underline{x}, \Delta \underline{x})}{\Delta x_1 \cdots \Delta x_N}, \text{ and}$$

$$H(S_\ell) = - \int_{\underline{x} \in S_\ell} \left\{ \frac{1}{M_\ell} \frac{\Delta n(\underline{x}, \Delta \underline{x})}{\Delta x_1 \cdots \Delta x_N} \log \left(\frac{1}{M_\ell} \frac{\Delta n(\underline{x}, \Delta \underline{x})}{\Delta x_1 \cdots \Delta x_N} \right) \right\} dx_1 \cdots dx_N.$$

With the weight vectors, $\underline{\omega}$, and the constraint, $\prod_{n=1}^N \omega_n = 1$, this entropy becomes:

$$\underline{H}_\omega(S_\ell) = \frac{1}{M_\ell \cdot \Delta V} \int_{\underline{x}' \in S_\ell} \Delta n(\underline{x}', \Delta \underline{x}') \log \left(\frac{M_\ell \Delta V}{\Delta n(\underline{x}', \Delta \underline{x}')} \right) dx'_1 \cdots dx'_N$$

Hence, $\underline{H}_\omega(S_\ell)$ attains a maximum when the distribution of $\Delta n(\underline{x}', \Delta \underline{x}')$ is as uniform as possible for varying $\underline{\omega}$.

The Algorithm OPTW-H:

1. Select some values $\Delta z_1, \Delta z_2, \dots, \Delta z_N$ with $\Delta z_n > 0$;
2. For each point $\underline{x} \in S$ simultaneously: take $\underline{x}$ as the center of a polycylinder, and expand $\underline{\Delta z}_n$ in each dimension $n, 1 \leq n \leq N$;

3. Repeat step 2 for all $\underline{x}, \underline{x}' \in S$ that have not yet reached the following stopping criterion; For some n , the expansion performed in step 2 leads to a nonempty intersection with the interior of an adjacent polycylinder with a point $\underline{x}' \in S$ as center. If this criterion is reached for some n , $\underline{x}$, and $\underline{x}'$, then the last expansion $\pm \Delta z_n$ in step 2 is deleted and further expansions along the n -th coordinate terminated.
4. If all expansions in step 2 are terminated by the criterion of step 3: For $1 \leq m \leq M$, the polycylinder around $\underline{x}_m \in S$ is given by its centerpoint, $\underline{x}_m$, and its radii $\Delta r_i^{(m)} = \Delta n_i^{(m)} \Delta z_i$, where $n_i^{(m)}$ is the number of expansions around $\underline{x}_m$ in the i -th dimension. Hence, the weight ω_n for the n -th coordinate is given by

$$\omega_n = \frac{1}{M} \sum_{m=1}^M n_i^{(m)} .$$

The result of this algorithm depends critically on the selection of $\Delta z_1, \dots, \Delta z_N$ in step 1, i.e. on the coarseness of the chosen raster.

2.2 The Clustering Approach to Scene Segmentation

The basic idea of clustering is to partition a set of objects into subsets such that each subset contains objects which are as similar to each other as possible. Each of these subsets is called a cluster (a review of clustering techniques developed before 1965 is given in [BAL65]). In order to investigate the applicability of clustering techniques to scene segmentation, the concept of clustering will be precisely stated and related to that of region finding in the next section. Subsequently,

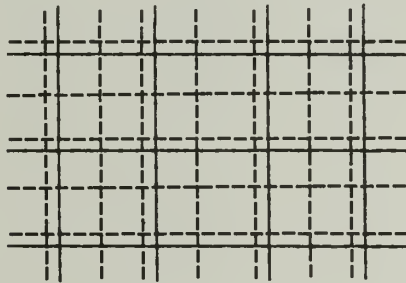
existing clustering techniques will be reviewed as far as they are used to obtain a clustering approach to scene segmentation.

2.2.1 A Formal Concept of Clustering

The underlying concept for the definition of a cluster, being developed in this section, is that a cluster, R , is a collection of objects such that the cohesion between all objects in R is somewhat larger than the sum of cohesions within the sets of any partition of S . A frequently used interpretation of cohesion is "average similarity". However, to apply clustering to region finding techniques, cohesion will be construed analogously to the mutual information defined in information theory.

Definition 2.15 For any standard tessellation, $T(\underline{x}^0)$, of the property space, X , a standard tessellation, $T'(\underline{x}^{0'})$, is called K -embedded in $T(\underline{x}^0)$ iff K is the smallest positive integer such that $\exists K$ translations $T_{\lambda_1}^{n_1} \dots T_{\lambda_K}^{n_K} \in T'$: $C = \bigcup_{k=1}^K T_{\lambda_k}^{n_k}(C')$

The following is an example for a $(4,4)$ -tessellation which is 4-embedded in another $(4,4)$ -tessellation:



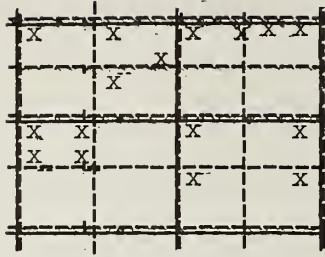
----- embedded tessellation

Definition 2.16 Let $T(\underline{x}^0)$ and $T'(\underline{x}^{0'})$ be standard tessellations of X and $T'(\underline{x}^{0'})$ be K -embedded in $T(\underline{x}^0)$. The cohesion $\zeta(R)$ of a set, R , consisting of P $T(\underline{x}^0)$ -cells, $C_1, \dots, C_P$ is defined by:

$$\zeta(R) = \sum_{p=1}^P H(C_p) - H(R) \quad .$$

Examples;

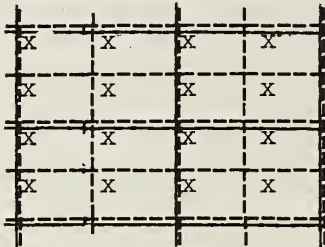
1.



— tessellation $T(\underline{x}^o)$
 - - - tessellation $T'(\underline{x}^o)$: each cell contains 4 points of X , 4-embedded in $T(\underline{x}^o)$.

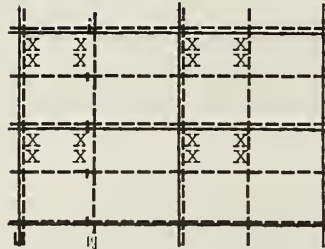
$P=4$, cohesion: 1.375 bits

2.



cohesion: 4.00 bits

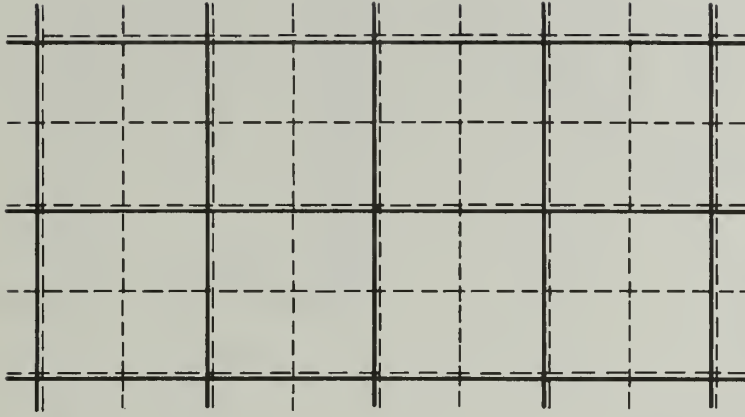
3.



cohesion: -2.00 bits

Cohesions for different scene segmentations are illustrated in Figures 2 and 3.

The cohesion, $\zeta(R)$, can be interpreted in terms of information theory as the average mutual information of the $T(\underline{x}^o)$ -cells constituting R on the ensemble of $T'(\underline{x}^o)$ -cells belonging to R . In the context of

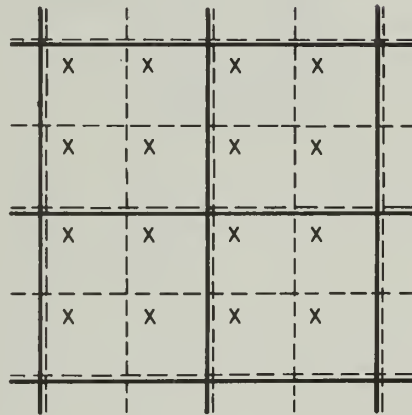


----- (4,4) - IMBEDDED TESSELLATION

Figure 2a (4,4) - Tessellation Which is
4-Imbedded in Another (4,4) - Tessellation

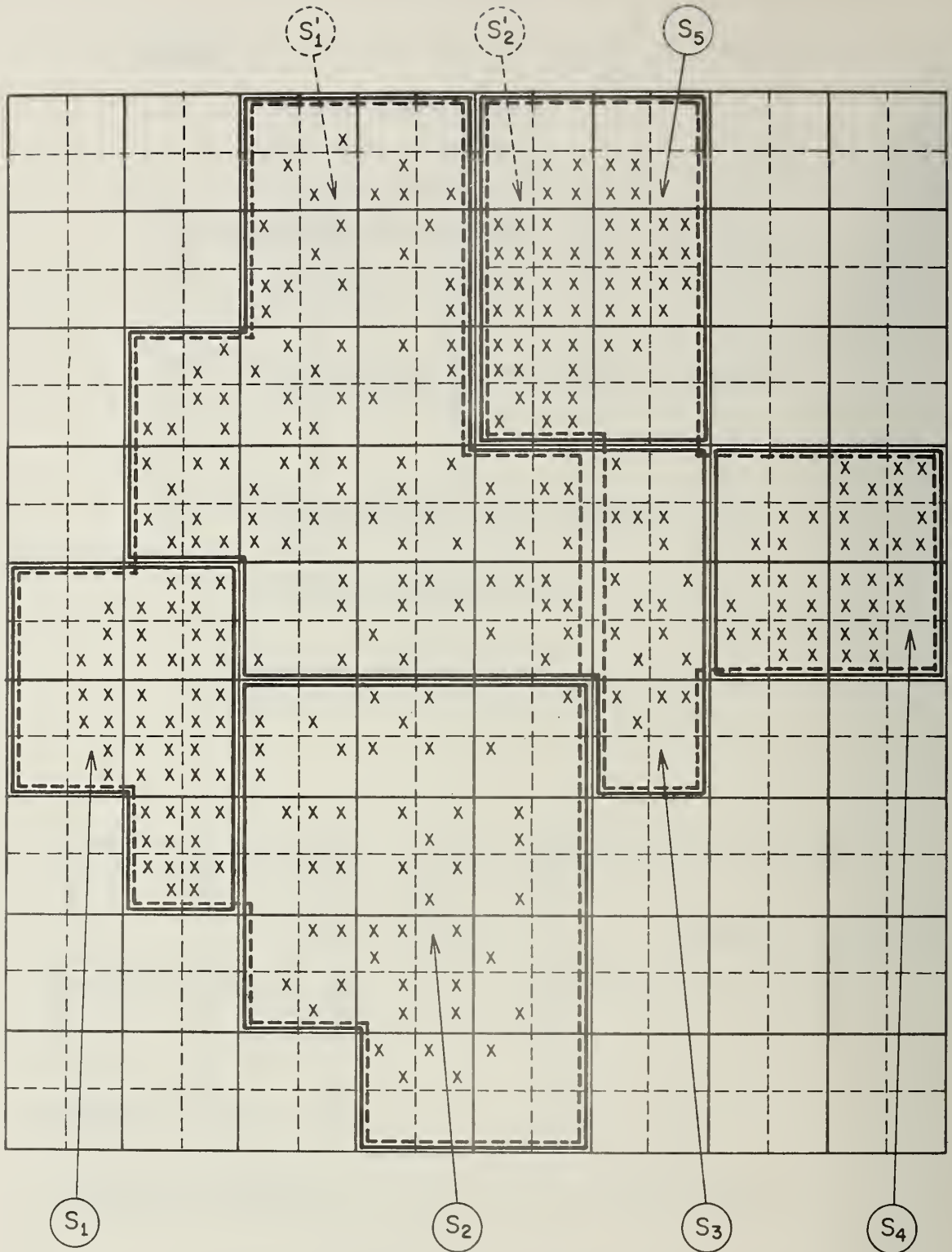


COHESION: 1.375 BITS



COHESION: 4.00 BITS

Figure 2b Cohesions of Simple Regions



COHESIONS: $\zeta(S_1) = 3.75$ BITS $\zeta(S_4) = 2.62$ BITS $\zeta(S'_1) = 1.94$ BITS
 $\zeta(S_2) = 3.78$ BITS $\zeta(S_5) = 3.75$ BITS $\zeta(S'_2) = 1.52$ BITS
 $\zeta(S_3) = 2.24$ BITS

Figure 3 Cohesions of Different Regions in a Scene

measure theory (cf. [HAL50]), it is shown in [WAT69a] that ζ is a supra-additive measure, i.e. for two sets, R_1 and R_2 , $\zeta(R_1 \cup R_2) \geq \zeta(R_1) + \zeta(R_2)$ if $\zeta(R_1) \cdot \zeta(R_2) > 0$. This is due to the fact that the entropy is a sub-additive measure, i.e. $H(R_1 \cup R_2) \leq H(R_1) + H(R_2)$. If the cohesion $\zeta(R)$ is negative, then for some i ($1 \leq i \leq P$), $H(C_i) = 0$. The extrema are given by $-\log P \leq \zeta(R) \leq P \log K - \log(P \cdot K)$. Definition 2.16 now permits us to define a cluster.

Definition 2.17 Let $T(\underline{x}^0)$ be a standard tessellation such that there exists another standard tessellation, $T'(\underline{x}^{0'})$, that is K -embedded in $T(\underline{x}^0)$ with $K > 1$. A chain-connected set, R , of P $T(\underline{x}^0)$ -cells is called an ϵ -cluster with respect to $T'(\underline{x}^{0'})$, if $\epsilon > 0$ is the largest number such that there exists a partition $\{R_1, \dots, R_S\}$ of R , $\zeta(R) - \sum_{p=1}^P \zeta(R_p) > \epsilon$.

2.2.2 Application of Clustering to Scene Segmentation

The task of finding an optimal partitioning of a picture into regions satisfying the [Region Criterion] can usually not be solved in a reasonable time by comparing all possible partitions. Instead, we employ the following strategy:

- (1) Using a parameterized heuristic rule, find disjoint chain-connected sets of picture points such that each set is part of exactly one of the regions to be found.
- (2) Each of the sets determined in (1) is used as a core for a region by applying a "grow algorithm" to each set so that eventually the entire picture is partitioned.
- (3) Changing the parameters of step (1) allows further repartitioning.

Psychologically, the decomposition of pictures into closed domains is usually guided by two aspects [ZUS70]; The attributes of points in a domain, (1) do not vary very much in their distribution, and (2) are quite similar to each other.

The first point of view led to the [Region Criterion] of Section 2.1.5. The second suggests the application of a clustering technique. To develop a method which employs both ideas, we make the following heuristic assumption [HA]: [HA] Each region of a satisfactorily decomposed picture contains exactly one distinguished cluster. A "distinguished cluster" in this context is an ϵ -cluster such that for any other ϵ' -cluster contained in the same region, $\epsilon \gg \epsilon'$.

Applying [HA] to the above strategy, we obtain the following algorithm for segmentation of a scene S :

Algorithm S: 1. Select some value ϵ ;
 2. Determine all ϵ -clusters in S , obtaining the clusters $S'_1, S'_2, \dots, S'_L$.
 3. Apply the algorithm GROW until all points in S are contained in one of the sets S_ℓ , $1 \leq \ell \leq L$, obtained by joining new points to S'_ℓ .

The algorithm GROW joins points to a previously established cluster center in accordance with the [Region Criterion]:

Algorithm GROW: Let $\underline{x} \in S$ be a point not yet contained in any of the sets $S_1, \dots, S_L$ of step 3 in the algorithm S. Perform the operation $S_\ell \leftarrow S_\ell \cup \{\underline{x}\}$ if
 (1) $S_\ell \cup \{\underline{x}\}$ is a chain-connected set and
 (2) $\forall \ell' \neq \ell; H(S_{\ell'} \cup \{\underline{x}\}) \leq H(S_\ell \cup \{\underline{x}\})$.

2.2.3 Review of Some Clustering Techniques

The necessity of using a clustering technique arises in step 2 of algorithm S. Although there have been many such techniques suggested (cf. [BAL65], [SS63]), their applicability in our case is severely restricted by:

1. The number of points to be clustered is very large ($>10^4$), but the number of attributes (gray value, color, etc.) is smaller than, for example, those considered in the numerical taxonomy of biological objects [SS63].
2. It is not possible to make a priori assumptions about the probability density distribution of points in the property space.
3. Similarity measures as well as the metric in the property space are subject to pre-clustering considerations and may be adapted during the clustering procedure.

Under these restrictions, we consider (1) a probabilistic and (2) a graph-theoretical technique. Their applicability to our scene segmentation approach will be evaluated in Section 2.2.4.

1. A Probabilistic Clustering Technique

We consider the probabilistic clustering technique of the following scheme [TSY71]:

For a given set, X , of patterns, $\underline{x}$, find L probability density distributions, $P_1, \dots, P_L$, such that the mixture density, $P(\underline{x}) = \sum_{\ell=1}^L p_{\ell} P_{\ell}(\underline{x}|\ell)$, attains each of its maxima ("modes") in exactly one of the disjoint, chain-connected subsets $X_1, \dots, X_L$ of X with $\bigcup_{\ell=1}^L X_{\ell} = X$.

This scheme satisfies restrictions 2. and 3. and is often referred to as an example for "self-learning" [TSY71]. In order to reduce the

computational complexity and in accordance with restriction 1. we reduce this scheme to estimating modes. After having found all modes, the remaining points in X can be classified according to the nearest neighbor classification rule which classifies a point, x , to the set containing its nearest neighbor. It is shown in [CH67] that the probability of error for this rule is less than twice the error-probability of Bayesian decision rules.

The following method for multivariant mode-seeking in a set, X , is proposed in [HF70] and [MD65]:

1. Compute the maximum eigen-vector (associated with the largest eigen-value), of the covariance matrix of X ;
2. Project X onto its maximum eigen-vector;
3. Determine the extrema of the one-dimensional probability density obtained in 2.;
4. Partition X with hyperplanes perpendicular to the eigen-vector found in step 1. and intersecting the eigen-vector at the locations of relative minima found in step 3.;
5. If only one extrema was found in step 3., the above procedure is repeated, starting at step 1. with the next largest eigen-vector;
6. For each new domain found in 4., the above procedure is repeated from step 1.

2. Graph-Theoretical Clustering Techniques

a. Matula's Clustering Concept

In [MAT70] and [MAT71], D.W. Matula introduced a concept of clustering which can be summarized in terms of our notation as follows:

A weighted graph, $G = (\Gamma, E)$, consists of a set, Γ , of nodes and a set, E , of edges between nodes in Γ such that each edge $e \in E$ is associated with a weight, $\omega(e)$. Let S be a set of points representing a picture in the property space, $X(A)$. S can be considered as a complete graph, $G(S) = (S, E(S)) = \{(\underline{x}, \underline{x}') \mid \underline{x}, \underline{x}' \in S\}$. We take the weight, $\omega(e)$, of an edge, $e = (\underline{x}, \underline{x}') \in E(S)$, to be the distance between its end points: $\omega(e) = \frac{1}{d(\underline{x}, \underline{x}')}.$ Consequently, the weight of an edge in the graph, $G(S)$, can be interpreted as the affinity between its endpoints. The affinity between two disjoint sets, $S', S'' \in S$ of nodes in $G(S)$ is given by

$$\omega(S', S'') = \frac{1}{|S'| |S''|} \sum_{\underline{x} \in S'} \sum_{\underline{x}' \in S''} \omega(\underline{x}, \underline{x}').$$

We consider $G(S)$ to be pruned by specifying the threshold affinity graph, $G_t = \{(\underline{x}, \underline{x}') \in E(S) \mid d(\underline{x}, \underline{x}') < t\}$ for some cut level $t > 0$.

If $\{G_1, G_2\}$ is a partition of any graph G , the cut set, $C(G_1, G_2)$, is given by $C(G_1, G_2) = \{(\underline{x}, \underline{x}') \mid \underline{x} \text{ is node in } G_1, \underline{x}' \text{ is node in } G_2\}$. A graph, G , of order ≥ 2 in which every cut set has $k \geq 0$ edges is called k-edge-connected, and a maximal k -edge-connected subgraph of G is called a k-component of G .

For any graph of order ≥ 2 , the edge-connectivity is defined as $\lambda(G) = \min\{|C| \mid C \text{ is a cut set of } G\}$. For any $x \in S \cup G(S)$, its cohesiveness $h(\underline{x})$ is defined as $h(\underline{x}) = \max\{\lambda(G') \mid G' \text{ is a subgraph of } G \text{ and } G' \text{ contains } \underline{x}\}$, and the strength of the graph, G , is $\sigma(G) = \max\{\lambda(G') \mid G' \text{ is a subgraph of } G\}$. Every k -component ($k \geq 1$) which does not contain a $(k+1)$ -component of G , as well as every trivial component (consisting of one vertex only) of G is a cluster of G . G itself is a cluster iff $\sigma(G) = \lambda(G)$. Any subgraph k' of a cluster k in G with $\lambda(k') = \lambda(k)$ is called a subcluster of G .

Under the relation "is a proper subgraph of", the subclusters of a graph form a partial order with clusters as maximal elements. An algorithm

determining the k-components and clusters of a graph is given in [MAT71].

b. Detection of Clusters Using Minimal Spanning Trees

In [ZAH71], an algorithm for detecting clusters using minimal spanning trees is given. The following summary is a reinterpretation of this concept adapted for application to our scene segmentation approach.

A spanning tree, $ST(S)$, of the graph, $G(S)$, defined under a. is a connected graph containing all nodes of $G(S)$ but no circuits. A minimal spanning tree, $MST(S)$ of $G(S)$ is a spanning tree with a minimal sum of all weights. If G_1 and G_2 are two interconnected subgraphs of G , their distance, $\rho(G_1, G_2)$, is defined as the minimal weight of all edges connecting G_1 and G_2 . The link set, $\lambda(G_1, G_2)$, is the set of all edges connecting G_1 and G_2 with weights $= \rho(G_1, G_2)$. Any subset, C , of S is called a δ -clump iff for any partition $\{C_1, C_2\}$ of C , $\rho(C, S-C) - \rho(C_1, C_2) \geq \delta$ with $\delta > 0$. It is shown in [ZAH71] that the restriction of an $MST(S)$ to a δ -clump in S is a connected subtree of the $MST(S)$.

A δ -clump is a set internally bound together stronger than the bounds between itself and nodes outside, whereas an ϵ -cluster of Definition 2.17 is completely defined by its own internal properties without respect to its environment.

An efficient algorithm for determining a minimum spanning tree of a graph is given in [SEP70] and others are referenced in [ZAH71]. The strategy to detect an ϵ -cluster, given in [ZAH71], is to determine inconsistent edges, whose weights differ significantly from the average weight of edges in the MST.

2.2.4 A Combined Clustering Technique for Scene Segmentation

When applying a graph clustering technique to a set, S , of points in

the property space, S is regarded as a complete graph, $G(S)$. Clusters, as defined in [MAT71], have to be obtained by taking appropriate values for t and searching for k -components in G_t for some k . [MAT70,71] do not give a way to choose t and k .

A connection between the concepts of [ZAH71] and [MAT70,71] is established by the following theorem:

- (*) Let G be a complete graph and $MST(G)$ be a minimal spanning tree of G . Then, the threshold affinity graph, G_t , partitions G into complete subgraphs, $G_1^t, G_2^t, \dots, G_{n(t)}^t$, iff there are exactly $n(t)$ edges, $(\underline{x}_1, \underline{x}'_1), \dots, (\underline{x}_{n(t)}, \underline{x}'_{n(t)})$, in $MST(G)$ with $d(\underline{x}_i, \underline{x}'_i) \geq t$ for $1 \leq i \leq n(t)$. The nodes of the subgraphs, $G_1^t, \dots, G_{n(t)}^t$, are those of the subtrees of $MST(G)$ after cutting all edges with weights $\geq t$.

Proof: This property follows immediately from Theorems 1-3 in [ZAH71].

Any edge in a minimal spanning tree connects two subtrees, and by Theorem 2 of [ZAH71], it is the edge with a smallest weight among all edges connecting the nodes of both subtrees in the complete graph containing all nodes of the tree. (*) allows us to reduce the problem of choosing the cut level, t , by considering all edges in the complete graph, G , to the task of looking at the edges of $MST(G)$ only.

The strength of a complete graph, G , with M nodes is $\mathcal{Q}(G) = \lambda(G) = M-1$, i.e. any complete graph is a cluster as defined in [MAT71] and has no subclusters except itself. Hence, the concept of clusters in [MAT71] is not applicable to our case. Instead, we apply the idea of cohesion developed in Section 2.2.1 to graphs.

Definition 2.18 Let T be a tree such that each edge, e , in T connecting two nodes, $\underline{x}$ and $\underline{x}'$, has a weight $\omega(e) = \frac{1}{d(\underline{x}, \underline{x}')}$. The entropy of T is defined as

$$H(T) = - \sum_{e \in T} \frac{\omega(e)}{\omega(T)} \log \frac{\omega(e)}{\omega(T)},$$

where $\omega(T) = \sum_{e \in T} \omega(e)$.

Definition 2.19 The cohesion, $\zeta(T; T_1, \dots, T_M)$, of a tree, T , with respect to a partition, $\{T_1, \dots, T_M\}$, of subtrees in T is defined as

$$\zeta(T; T_1, \dots, T_M) = \sum_{m=1}^M H(T_m) - H(T).$$

With Definitions 2.18 and 2.19, ϵ -clusters are also defined for trees.

An efficient clustering algorithm has to cope with the problem of finding appropriate partitions, $\{T_1, \dots, T_M\}$, of subtrees. The method of detecting inconsistent edges in [ZAH71] assumes a priori knowledge about the distribution of weights, which was excluded by restriction 2. of Section 2.2.3. An algorithm leading to quite optimal results is given in [WAT69a] but requires comparison of all possible dichotomies when sequentially partitioning the graph, thus violating restriction 1. Therefore, it seems appropriate to apply a probabilistic method which is more efficient when dealing with a large number of objects. This can be done by virtue of the following definition:

Definition 2.20 Let G be a graph and $\underline{x}$ a node in G . Then the weight, $\omega(\underline{x})$, is defined as

$$\omega(\underline{x}) = \frac{1}{|E(\underline{x})|} \sum_{e \in E(\underline{x})} \omega(e), \quad \text{if } |E(\underline{x})| \geq 2,$$

where $E(\underline{x})$ is the set of all edges having $\underline{x}$ as one endpoint and $\omega(e) = \frac{1}{d(\underline{x}, \underline{x}')}$, if $\underline{x}'$ is the other endpoint of e .

For $|E(\underline{x})| = 1$, we define $\omega(\underline{x}) = 0$.

If $\omega(G) = \sum_{\underline{x} \text{ node in } G} \omega(\underline{x})$, the normalized weight of a node is denoted by

$$\hat{\omega}_G(\underline{x}) = \frac{\omega(\underline{x})}{\omega(G)}.$$

Interpreting the normalized weight of a node as a probability density, we can now apply the mode seeking algorithm indicated in Section 2.2.3 to minimal spanning trees. The resulting modes are nodes in the MST to be used as cluster centers. The distance between two nodes in a tree is given by the sum of all weights $\omega(e)$ of edges forming the connecting path between the nodes. After having determined the modes of a MST, a partition of subtrees can be obtained by applying the nearest neighbor classification technique.

Combining the preceding concepts, we obtain the following algorithm to partition a chain-connected set, S , of points in the property space into regions;

Algorithm C

1. In the complete graph, $G(S)$, determine the minimal spanning tree, $MST(S)$, using the distances in the property space as weights of edges in $G(S)$;
2. Compute normalized weights of all nodes in $MST(S)$;
3. Apply the mode-seeking algorithm of Section 2.2.3 to $MST(S)$, yielding modes $\underline{x}_1^C, \underline{x}_2^C, \dots, \underline{x}_n^C$, defined as the cluster modes in $MST(S)$;
4. Using $\{\underline{x}_1^C\}, \dots, \{\underline{x}_n^C\}$ as initial classes, apply the nearest neighbor classification method to obtain a partition, $\{S_1, \dots, S_n\}$ of chain-connected sets in S ;
5. Using parameters θ_1 and θ_2 , repartition $\{S_1, \dots, S_n\}$ by applying the algorithms MERGE and SLICE.
 - a. Algorithm MERGE: Merge two adjacent subtrees, S_i and S_j of $MST(S)$ if $\zeta(S_i \cup S_j, S_i, S_j) > \theta_1$, for some parameter θ_1 .
 - b. Algorithm SLICE: Partition a subtree, S_i , of $MST(S)$ into subtrees, $S_{i1}, S_{i2}, \dots, S_{in(i)}$, if applying steps 1 to 4 of Algorithm C yields a partition $\{S_{i1}, \dots, S_{in(i)}\}$ of subtrees in S_i with $\zeta(S_i; S_{i1}, \dots, S_{in(i)}) > \theta_2$.

CHAPTER 3

STRUCTURAL ANALYSIS OF SCENES

3.1 Data Representation of Composite Scenes

We assume to be given the following results of the scene segmentation procedure described in Chapter 2:

- I. A scene, S , has been partitioned into M regions, $S_1, \dots, S_M$;
- II. Considering a set $A^r = \{A_1^r, \dots, A_T^r\}$ of T attributes applicable to regions, the vector $(A_1^r(S_m), \dots, A_T^r(S_m))$ has been evaluated for each region, S_m , in S .
- III. Let $R^r = \{R_1^r, \dots, R_S^r\}$ be a set of functions

$$R_s^r: \begin{cases} S \times S \rightarrow W_s^r = W_s^{r'} \cup \{*\} & \text{with } 1 \leq s \leq S, \\ (S_i, S_j) \rightarrow R^r(S_i, S_j) & 1 \leq i \neq j \leq M \end{cases}$$

mapping pairs of different regions into a value set, W_s^r , which consists of a discrete finite set, $W_s^{r'}$ of real numbers and the don't care value, $*$.

For each pair, (S_i, S_j) , of different regions, the vector $(R_1^r(S_i, S_j), \dots, R_S^r(S_i, S_j))$ has been evaluated.

- Remarks:
- 1. The attributes $A_1^r, \dots, A_T^r$ map regions into discrete, finite value sets, V_t^r ($1 \leq t \leq T$), as do the local attributes introduced in Definition 2.1. Two different types of such attributes can be considered:
 - a. attributes obtained by averaging local attributes over all picture points contained in a region. (Example: The arithmetic mean of gray values of points in a region.)

(Example: An attribute indicating the shape of a region, e.g. with a value set $\{1, 2, 3, *\}$, where $1 = \text{circular}$, $2 = \text{triangular}$, $3 = \text{rectangular}$, and $* = (\text{not } 1, 2, \text{ or } 3 \text{ or not specified}).$)

2. The functions, $R_1^r, \dots, R_S^r$ can be interpreted as binary relations with many values. (Example: The quantized distance, $d^r(S_i, S_j)$, of two regions, S_i and S_j is

$$d^r(S_i, S_j) = \left[\frac{1}{|S_i| |S_j|} \sum_{\underline{x}_i \in S_i} \sum_{\underline{x}_j \in S_j} ||\underline{x}_i - \underline{x}_j|| \right],$$

where $||\underline{x}||$ denotes the Euclidean norm of a point, $\underline{x} = (x_1, x_2)$, in the plane and $[r]$ denotes the smallest integer less than or equal to r .)

3. Although n -ary relations with $n > 2$ may be used to describe properties of regions in a scene, the restriction to binary-relations, like the functions R_S^r , does not restrict the generality of scene description. It is shown in [MON71] that n -ary relations with $n > 2$ can be reasonably well represented or approximated by binary relations.

We consider two representations of scenes:

- A. The Relationship Matrix - The relationship matrix, $R = [r_{ij}]$, is introduced in [CR71]. For a scene partitioned into M regions, $S_1, \dots, S_M$, R is an $M \times M$ matrix. The entries, r_{ij} , of R are:

for $i \neq j$: S -dimensional vectors, $(R_1^r(S_i, S_j), \dots, R_S^r(S_i, S_j))$

for $i = j$: T -dimensional vectors, $(A_1^r(S_i, S_j), \dots, A_T^r(S_i, S_j))$

A complete description of all properties of a region, S_i , is provided by the i -th row and the i -th column of R . R is symmetric if all functions, $R_s^r \in R^r$, are symmetric with respect to their arguments, $(S_i, S_j) \in S^2$.

B. The Digraph Representation - A scene, $S = \{S_1, \dots, S_M\}$, can be represented as a digraph, $G^r = (\Gamma^r, E^r)$, whose node set is $\Gamma^r = \{S_1, \dots, S_M\}$. The set, E^r , of edges in G^r consists of labelled and weighted edges. There is an edge, e_{ij}^u , such that:

for $i \neq j$: u , with $1 \leq u \leq S$, indicates the subscript in R_u^r of a function in R^r and e_{ij}^u is associated with the weight, $\omega(e_{ij}^u) = R_u^r(S_i, S_j)$.

for $i = j$: u , with $1 \leq u \leq T$, indicates the subscript of an attribute in A^r and e_{ij}^u is associated with the weight, $\omega(e_{ij}^u) = A_u^r(S_i)$.

Remark: To simplify this representation, an edge, (S_i, S_j) , with a label, u , is deleted whenever $R_u^r(S_i, S_j) = *$.

3.2 Synthesis of Composite Objects in Scenes

The scene segmentation procedure described in Section 2 results in partitioning scenes into primitive regions. Usually, scenes consist of composites of primitive regions, called figures. Hence, scene analysis has to cope with synthesizing figures from primitive regions. The synthesis of figures is done by merging regions according to some criteria to be established. However, these criteria depend upon the specific environment and the intentions considered. In this section, the formal structure of

two such approaches is briefly reviewed, using the scene representations introduced in Section 3.1.

3.2.1 Synthesis of Composites Using Graph Transformation Rules

Let $G^r = (\Gamma^r, E^r)$ be a digraph representing the scene $S = \{S_1, \dots, S_n\}$.

The synthesis of composites is obtained by merging regions, yielding a new graph $G_c^r = (\Gamma_c^r, E_c^r)$. Nodes in Γ_c^r may either represent primitive regions in S or regions obtained by merging regions in S . The edges between any two nodes in Γ_c^r , which are also both nodes in Γ^r , remain unchanged. All other edges in E_c^r have to be determined. The function mapping G^r to G_c^r is a special graph transformation.

It is shown in [SMC 70] that many graph transformation rules can be conveniently inferred from invariance principles such as conservation of union and intersection of point sets.

3.2.2 Synthesis of Composites Using Cartesian Covers

The concept of cartesian covers is introduced in [MMC 70]. A discrete, n -dimensional set $E = \{(x_1, \dots, x_n) \mid x_i \in H_i \text{ for } 1 \leq i \leq n\}$ with $H_i = \{0, 1, \dots, h_i - 1 \mid h_i > 0\}$ ($1 \leq i \leq n$) is considered. An element $(x_1, \dots, x_n)$ in E is usually denoted by e^j with

$$j = x_n + \sum_{i=1}^{n-1} (x_{n-i} \prod_{k=0}^{i-1} h_{n-k})$$

and $e^j \in E$ is called an event.

A cartesian literal, X_i^{Ai} , is defined as the set

$$X_i^{Ai} = \{(x_1, \dots, x_n) \mid x_i \in A_i\}$$

for $A_i \subseteq H_i$ ($1 \leq i \leq n$).

A set $L \subseteq E$ is called a cartesian complex if it is represented as the intersection of cartesian literals: $L = \bigcap_{i \in I} X_i^{A_i}$, where $I \subseteq \{1, \dots, n\}$.

L is also called

an interval, if there are n numbers $a_i, b_i \in H_i$ ($i=1, \dots, n$) such that

$$A_i = \{x_i \mid a_i \leq x_i \leq b_i\} \text{ for all } i, 1 \leq i \leq n;$$

a factor, if there do not exist $a_i, b_i \in H_i$ (for all $i, 1 \leq i \leq n$) such that $A_i = \{x_i \mid a_i \leq x_i \leq b_i\}$.

Example: Let $n = 2, h_1 = h_2 = 3; A_1 = \{1, 2\}, A_2 = \{0, 1\}, A'_1 = \{0, 2\},$

and $A'_2 = \{0, 2\};$

then, $L = X_1^{A_1} \cap X_2^{A_2}$ is an interval, while $L' = X_1^{A'_1} \cap X_2^{A'_2}$ is

a factor.

Let $f: E \rightarrow \{[0, 1], *\}$ be a mapping defining the sets $F^\alpha = \{e \in E \mid f(e) = \alpha\}$

for $\alpha \in \{0, 1, *\}$, $F^\phi = \{e \in E \mid 0 < f(e) < 1\}$, and for some $\lambda \in [0, 1]$:

$$F^{1\lambda} = \{e \in E \mid f(e) \geq \lambda\}, F^{0\lambda} = \{e \in E \mid f(e) < \lambda\}. \text{ A set } D(f|\lambda) = \{L_i \mid i=1, 2, \dots\}$$

of cartesian complexes is called a cartesian cover of f under λ if

$F^{1\lambda} \subseteq \bigcup_i L_i \subseteq F^{1\lambda} \cup F^*$. If all cartesian complexes in $D(f|\lambda)$ are intervals (factors), $D(f|\lambda)$ is also called an interval (factor) cover. For simplicity, we assume $\lambda = 1$ and write $D(f|\lambda) = D(f)$, $F^{1\lambda} = F^1$, and $F^{0\lambda} = F^0$.

The application of cartesian covers to the formation of composites consists of two steps:

1. We interpret the property space as an n -dimensional discrete vector space E . Then, the attribute values of a region are represented by an event in E . Thus, a scene $S = \{S_1, \dots, S_n\}$ is represented in E by n events $e_1, \dots, e_n$.

2. We establish criteria to define the optimality of a cartesian cover. Then, we try to find an optimal set of cartesian complexes covering the events $e_1, \dots, e_n$. A composite is formed by joining all events covered

by exactly one cartesian complex.

An efficient algorithm to form optimal cartesian covers is described in [MR 72]. Examples for criteria defining the optimality of a cartesian cover L are:

- a) $E(L \cap F^1) = \text{Max.}$ - maximize the number of events in F^1 covered by complexes in L ;
- b) $\ell(L) = \text{Min.}$ - minimize the number of literals needed to represent L ;
- c) $\text{den}(L, F^1) = \text{Max.}$ - maximize $E(L \cap F^1)$ per literal in L .

where

$$\text{den}(L, F^1) := \frac{E(L \cap F^1)}{\ell(L)}$$

For synthesizing switching circuits, the importance of these criteria is well established. However, in the case of scene analysis, little is known whether these criteria do actually contribute to the description of regions, and further experimentation is necessary. For this purpose, weights for specific attributes or events may be introduced to compute weighted covers.

3.2.3 Matching Models to Scenes

It seems to be appropriate to assume that most scenes we are dealing with are composed of composite figures such that each can be regarded as a realization of some known 'model figure'. Instead of imposing criteria of hardly known importance and meaning, we may try to fit models to a scene. For this purpose, the relationship matrix appears to be a useful data structure.

A model, $\underline{M}$, is a collection of m objects $\underline{M} = \{M_1, \dots, M_m\}$, represented by an (m, n) -relationship matrix $R(\underline{M})$. We assume that we are given

a set $\{\underline{M}_1, \dots, \underline{M}_p\}$ of p models. Then, the task of interpreting a scene $S = \{S_1, \dots, S_n\}$ in terms of models, $\underline{M}_1, \dots, \underline{M}_p$, can be defined as finding an optimal match of S to some $\underline{M}_j$ such that (a) each region in S is associated with exactly one model, and (b) the number of regions not or only poorly matching a model is kept at a minimum. Although approaches toward matching algorithms have been made (e.g. [BP 70]), their efficiency is still rather poor.

CHAPTER 4

CONCLUSIONS

The property space representation allows us to define the density of the spatial distribution of local properties of picture points. Properly normalized, this density can be interpreted as a probability density which serves to define entropy and cohesion of a set of picture points. It is shown for an arbitrary but simple example that the cohesion gives a good theoretical basis for applying clustering methods. Clearly, much more experimentation is necessary to establish this concept.

Similar to parsing in syntax analysis, scene analysis has (1) to infer a global structure from local properties, and (2) to recognize given structures by matching their components to regions in the picture. Successful parsers usually employ a combination of the 'top-down' and 'bottom-up' method, and this strongly suggests an approach combining these two schemes in scene analysis, too. Therefore, it seems to be highly desirable to extend global analysis techniques as graph transformations and covering so that they will also employ local properties.

REFERENCES

- [ATT 54] Attneave, F., "Some Informational Aspects of Visual Perception", *Psychology Review*, vol. 61 (1954), pp. 183-193.
- [AA56] Attneave, F. and Arnoult, M.D., "The Quantitative Study of Shape and Pattern Perception", *Psychology Bulletin*, vol. 53 (1956), pp. 452-471.
- [BAL 65] Ball, G.H., "Data Analysis in the Social Sciences - What About the Details?", *Proceedings of the 1965 Fall Joint Computer Conference*, vol. 27, part 1 (1965), pp. 533-559.
- [BP 70] Barrow, H.G. and Popplestone, R.J., "Relational Descriptions in Picture Processing", in *Machine Intelligence, Vol. 6*, Melzer, B. and Michie, D. (eds.), Edinburgh University Press, 1970, pp. 377-396.
- [BIR 67] Birkhoff, G., *Lattice Theory*, American Mathematical Society, Providence, Rhode Island, 1967.
- [BON 64] Bonner, R.E., "On Some Clustering Techniques", *IBM Journal of Research and Development*, January 1964, pp. 22-32.
- [BF 70] Brice, C.R. and Fennema, C.L., "Scene Analysis Using Regions", *Artificial Intelligence*, vol. 1 (1970), pp. 205-226.
- [CH 67] Cover, T.M. and Hart, P.E., "Nearest Neighbor Pattern Classification", *IEEE Transactions*, vol. IT-13 (1967), pp. 21-27.
- [CR 71] Chien, Y.T. and Ribak, R., "Relationship Matrix as a Multi-Dimensional Data Base for Syntactic Pattern Generation and Recognition", *Proceedings of the Two Dimensional Digital Signal Processing Conference*, October 1971, Columbia, Missouri.
- [CUL 68] Cullen, H.F., *Introduction to General Topology*, Boston, 1968.
- [FER 67] Ferguson, T.S., *Mathematical Statistics*, Academic Press, New York, 1967.
- [FOY 64] Foy, W.H., "Entropy of Simple Line Drawings", *IEEE Transactions*, vol. IT-10, April 1964, pp. 165-167.
- [FRA 67] Fralick, S.C., "Learning to Recognize Patterns Without a Teacher", *IEEE Transactions*, vol. IT-13, January 1967, pp. 57-64.

- [FU 69] Fu, K.S., "On Sequential Pattern Recognition Systems", in Methologies of Pattern Recognition, Watanabe, S. (ed.), Academic Press, New York, 1969.
- [GAL 68] Gallager, R.G., Information Theory and Reliable Communication, Wiley and Sons, New York, 1968.
- [GC 66] Green, R.T. and Courtis, M.C., "Information Theory and Figure Perception: The Metaphor that Failed", *Acta Psychologica*, vol. 25 (1966), pp. 12-35.
- [GR 69] Gower, J.C. and Ross, G.J.S., "Minimum Spanning Trees and Linkage Cluster Analysis", *Applied Statistics*, vol. 18 (1969), pp. 54-64.
- [GRA 71] Gray, S.B., "Local Properties of Binary Images in Two Dimensions", *IEEE Transactions*, vol. C-20 (1971), pp. 551-561.
- [GRE 65] Grenander, U., "Some Direct Estimates of the Mode", *Ann. Math. Stat.*, vol. 36 (1965), pp. 131-138.
- [GUZ 68] Guzman, A., "Decomposition of Scenes into Bodies", *AFIPS Conference Proceedings*, vol. 33, part 1 (1968), pp. 291-304.
- [GUZ 70] Guzman, A., "Analysis of Curved Line Drawings Using Context and Global Information", in Machine Intelligence, Vol. 6, Meltzer, B. and Michie, D. (eds.), Edinburgh University, 1970, pp. 325-375.
- [HAL 50] Halmos, P.R., Measure Theory, Princeton Press, 1950.
- [HA 67] Hart, P.E., "A Brief Survey of Preprocessing for Pattern Recognition", Stanford Research Inst. Report No. RADC-TR-66-819 (1967).
- [HA 69] Harary, F., Graph Theory, Addison-Wesley, New York, 1969.
- [HARL 70] Harlick, R.M., "Multi-Image Clustering", *Proceedings of the 1970 Army Numerical Analysis Conference*, pp. 75-90.
- [HF 70] Hennchon, E.G. and Fu, K.S., On Mode Estimation in Pattern Recognition, Purdue University Press, Lafayette, Indiana, 1970.
- [HK 67] Haralick, R.M. and Kelly, G.L., "Pattern Recognition with Measurement Space and Apatial Clustering for Multiple Images", *Proceedings of the IEEE*, vol. 57, April 1967.
- [HY 61] Hocking, J.G. and Young, G.S., Topology, Addison-Wesley, New York, 1961.

- [KIT 63] Kittel, C., Introduction to Solid State Physics, 2nd Edition, John Wesley, New York, 1963.

- [MAT 70] Matula, D.W., "Cluster Analysis via Graph Theoretic Techniques", Proceedings of the Louisiana Conference on Combinatorics, Graph Theory, and Computing, Mullin, R.C. (ed.), University of Manitoba, 1971, pp. 199-212.

- [MAT 71] Matula, D.W., "k-Components, Clusters, and Slicings in Graphs", preprint in 1971. To appear in SIAM Journal of Applied Mathematics.

- [MD 65] Mattson, R.L. and Dammann, J.E., "A Technique for Determining and Coding Subclasses in Pattern Recognition Problems", IBM Journal of Research and Development, July 1965.

- [MI 72] Michalski, R.S., "Varivalued Logic and Its Applications to Pattern Recognition", Department of Computer Science, University of Illinois, Urbana, Illinois Report, 1972 (in preparation).

- [MMC 72] McCormick, B.H. and Michalski, R.S., "CARTESIAN COVERS - A Theoretical Introduction", Department of Computer Science, University of Illinois, Urbana, Illinois, 1972 (in preparation).

- [MON 71] Montanari, U., "Networks of Constraints: Fundamental Properties and Applications to Picture Processing", Department of Computer Science Report, Carnegie-Mellon University, Pittsburgh, Pennsylvania, January 1971.

- [MP 67] Minsky, M.A. and Papert, S., Project MAC Progress Report IV, MIT Press, Cambridge, Massachusetts, 1967.

- [MR 72] Michalski, R.S. and Raulefs, P., "Computer Synthesis of Cartesian Covers and Varivalued Logic Expressions", Department of Computer Science Report, University of Illinois, Urbana, Illinois, 1972 (in preparation).

- [ORE 62] Ore, O., Theory of Graphs, American Mathematical Society Publication, vol. 38, Providence, Rhode Island, 1962.

- [PR 70] Preparata F.P. and Ray, S.R., "An Approach to Artificial Non-symbolic Cognition", Coordinated Science Laboratory Report No. R-478, University of Illinois, Urbana, Illinois, 1970.

- [ROB 65] Roberts, L.G., "Machine Perception of Three Dimensional Solids", Optical and Electro-Optical Information Processing, Tippet, et. al. (eds.), MIT Press, Cambridge, Massachusetts, 1965.
- [ROS 69] Rosenfeld, A., Picture Processing by Computer, Academic Press, New York, 1969.
- [SEB 62] Sebestyen, G.S., Decesion-Making Processes in Pattern Recognition, New York, 1962.
- [SEP 70] Seppanen, J.J., "Algorithm 399", Communications of the ACM (Oct. 1970), pp. 621-622.
- [SMC 70] Schwebel, J.C. and McCormick, B.H., "Consistent Properties of Composite Formation Under a Binary Relation", Information Sciences, vol. 2 (1970), pp. 179-209.
- [SS 63] Sokal, R.R. and Sneath, P.H.A., Principles of Numerical Taxonomy, San Francisco, 1963.
- [TOU 69] Tou, J.T., "Feature Selection for Pattern Recognition Systems", in Methodologies of Pattern Recognition, Watanabe, S. (ed.), Academic Press, New York, 1969, pp. 493-508.
- [TSY 71] Tsypkin, Ya. Z., Adaption and Learning in Automatic Systems, Academic Press, New York, 1971.
- [WAT 60] Watanabe, S., "Information-Theoetical Aspects of Inductive and Deductive Inference", IBM Journal of Research and Development, vol. 4 (1960), pp. 208-231.
- [WAT 69a] Watanabe, S., Knowing and Guessing, New York, 1969.
- [WAT 69b] Watanabe, S., "Pattern Recognition as an Inductive Process", Methodologies of Pattern Recognition, Watanabe, S. (ed.), Academic Press, New York, 1969, pp. 521-534.
- [ZAH 71] Zahn, C.T., "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters", IEEE, vol. C-20 (1971), pp. 66-86.
- [ZUS 70] Zusne, L., Visual Perception of Form, Academic Press, New York, 1970.

U. S. ATOMIC ENERGY COMMISSION
UNIVERSITY-TYPE CONTRACTOR'S RECOMMENDATION FOR
DISPOSITION OF SCIENTIFIC AND TECHNICAL DOCUMENT

(See Instructions on Reverse Side)

1. AEC REPORT NO. COO-2118-0028 UIUCDCS-R-72-496	2. TITLE METHODOLOGICAL ASPECTS OF SCENE SEGMENTATION
--	--

3. TYPE OF DOCUMENT (Check one):

- ☒ a. Scientific and technical report
- ☐ b. Conference paper not to be published in a journal:
Title of conference _____
Date of conference _____
Exact location of conference _____
Sponsoring organization _____
- ☐ c. Other (Specify) _____

4. RECOMMENDED ANNOUNCEMENT AND DISTRIBUTION (Check one):

- ☒ a. AEC's normal announcement and distribution procedures may be followed.
- ☐ b. Make available only within AEC and to AEC contractors and other U.S. Government agencies and their contractors.
- ☐ c. Make no announcement or distribution.

5. REASON FOR RECOMMENDED RESTRICTIONS:

6. SUBMITTED BY: NAME AND POSITION (Please print or type)

Peter Raulefs
Fellow, Department of Computer Science

Organization

Department of Computer Science
University of Illinois
Urbana, Illinois 61801

Signature

Peter Raulefs

Date

January 24, 1972

FOR AEC USE ONLY

7. AEC CONTRACT ADMINISTRATOR'S COMMENTS, IF ANY, ON ABOVE ANNOUNCEMENT AND DISTRIBUTION RECOMMENDATION:

PATENT CLEARANCE:

- ☐ a. AEC patent clearance has been granted by responsible AEC patent group.
- ☐ b. Report has been sent to responsible AEC patent group for clearance.
- ☐ c. Patent clearance not required.

BIBLIOGRAPHIC DATA HEET	1. Report No. UIUCDCS-R-72-496	2.	3. Recipient's Accession No.
	Title and Subtitle Methodological Aspects of Scene Segmentation		5. Report Date June, 1972
Author(s) Peter Raulefs		6.	
Performing Organization Name and Address Dept. of Computer Science Univ. of Illinois at Urbana-Champaign Urbana, Illinois 61801		8. Performing Organization Rept. No.	10. Project/Task/Work Unit No. ILLIAC III
Sponsoring Organization Name and Address U. S. Atomic Energy Commission		11. Contract/Grant No. AT(11-1)-2118	13. Type of Report & Period Covered
Supplementary Notes		14. Feb. - June 1972	
Abstracts Scene segmentation into regions and structural analysis of scenes is discussed. The spatial distribution of properties of points in a digitized picture is used to define the entropy of point sets (regions) in a picture. An optimal partitioning of a scene into regions is given by maximizing the average entropy of all regions. Using entropy, the cohesion of regions is defined and applied to analyzing the clustering approach to scene segmentation. Structural analysis of scenes in terms of synthesizing composites is briefly reviewed by considering graph transformations and cartesian covers.			
Key Words and Document Analysis. 17a. Descriptors			
Identifiers/Open-Ended Terms			
COSATI Field/Group			
Availability Statement RELEASE UNLIMITED		19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages
		20. Security Class (This Page) UNCLASSIFIED	22. Price

SEP 21 1972

JUL 26 1973



UNIVERSITY OF ILLINOIS-URBANA

510.84 IL6R no. C002 no.493-498(1972
Tree height reduction for parallel proce



3 0112 088400210